DOCUMENT RESUME

ED 140 357 CS 203 518

AUTHOR Tanner, Linda R.
TITLE Some Constraints on the Word Frequency Effect in Written Discourse.
PUB DATE 76
NOTE 22p.; Report prepared at University of Minnesota

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Content Analysis; Grade 6; Intermediate Grades; *Reading Comprehension; *Reading Research; *Sex Differences; Teaching Techniques; *Vocabulary; *Word Frequency

ABSTRACT

This study investigates the effect of word frequency on reading comprehension, in relation to type of content, vocabulary instruction method, student sex, and question type. Subjects, 120 sixth-grade readers, responded to three types of comprehension questions after reading two types of stories. Each subject had previously received one of three vocabulary treatments. Analysis of variance revealed that, while alteration of word frequency did not affect comprehension, alteration of question and content type did significantly affect comprehension. Question type and content type were found to interact with each other and with other factors, indicating the need for controlling these two factors in subsequent comprehension research. (Author/KS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED140357

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING T POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

#5

SOME CONSTRAINTS ON THE WORD FREQUENCY EFFECT IN WRITTEN DISCOURSE

Linda R. Tanner

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Linda R. Tanner

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

CS 203 518

The effects of word frequency on comprehension were investigated in relation to content type, vocabulary instruction, student sex and question type. 120 sixth grade (competent) readers responded to three types of comprehension questions after reading two types of stories and receiving one of three vocabulary treatments. Analysis of variance revealed that altering word frequency did not affect comprehension, but altering question type and content type did significantly affect comprehension scores. Question type and content type interacted each other and with other factors, indicating the necessity of controlling these two factors in comprehension research.

Word frequency has long been an influential variable in explaining and predicting the difficulty people have in processing written material. Its effect on visual duration threshold is a well-replicated finding (e.g. Solomon and Postman, 1952; Howes and Solomon, 1951): more frequent words are recognized at lower thresholds; its effect on children's word identification ability is a more recent finding (e.g. Pearson and Studt (1975)). Historically, it has been included as a direct (e.g. the Dale-Chall formula (1948)) or an indirect (e.g. using word length as a surrogate as in Fry (1968)) factor in readability formulas, even though Klare (1974-75) reports that direct manipulation of frequency has only occasionally produced effects on comprehension.

However, in direct conflict to Klare's conclusion is the finding of Marks, Doctorow and Wittrock (1974) that altering the word frequency values of one-sixth of the running words in story length passages had a significant effect on the comprehension of sixth-grade students irrespective of reading ability.

What Marks, Doctorow and Wittrock did was to divide the passages into blocks of six words. Then for one-half of the blocks, they located equally

long (same number of letters) lower frequency synonyms for higher frequency words in the text. In the other half, they located higher frequency synonyms for lower frequency words. The net result was the creation of two versions of each passage that differed only in the frequency values of one-sixth of the words. Word length was constant across versions. By testing a mixture of literal and inferential multiple-choice questions after each version, they determined that manipulating word frequency substantially altered comprehension when readability was held constant. They concluded that authors of text materials should heed their findings and consider the effect they might have if they searched for simpler synonyms for some of their more baffling lexical entries.

Several questions arise from their study. First, can one locate words of equal length but different frequencies that are truly synonomous? It is possible that conceptual difficulty was not maintained between versions of a passage. We did not find this to be true, however. We asked a group of ten graduate students to judge the synonymity of the synonym pairs used by Marks et al. While the pairs were regarded as unacceptable out of context, they were judged acceptable in context.

Second, did they really hold readability constant? On the face of it, any readability formula that uses word length as a factor in its regression equation would yield a positive answer. However, one has to ask why word length is used as a factor in a readability formula. Historically, it arose as an extrapolation from Zipf's (1945) Law that there is an inverse logarithmic relationship between word frequency and word length: less frequent words tend to be longer than more frequent words. If word frequency and word length were plotted on a graph using arithmetic (rather than logarithmic) increments on both axes, the scatter-plot would resemble a fat cigar more than a straight line. Hence it would

be (and is) possible to find synonyms of variant frequencies that do not differ in length, even though, in general, frequency values decrease as length increases. In short, it is possible that Marks, Doctorow, and Witrock were taking advantage of the imperfect nature of the relationship between frequency and length. By this argument, a readability formula that uses a direct measure of frequency ought to yield variant readability values for their passages, which were equated for length but varied in word frequency. This is exactly what we found when we applied the Dale-Chall (1948) formula to their passages. For example their passage, "Conductor Moses" was rated as 6.9 in the high frequency version but 10.5 in the low frequency version. Likewise, "Bail Out" was rated 7.5 in the high frequency version and 10.2 in the low frequency version. Over all five passages used by Marks, Doctorow and Wittrock, the typical difference between high and low frequency versions was about three grade levels.

Even so, the question of the influence of the frequency affect remains. All we have done is to show that readability was confounded in their experiment. The effect of word frequency is an important educational issue. If one recommends that high frequency words be used whenever they fit the content, one very quickly gets into a logical paradox and an instructional bind. By the logic of the high frequency rule, one would ultimately end up with a few well-known words to communicate the range of complexity of human experience. The instructional bind is, of course, How are students ever to learn any new words if they are exposed only to familiar words.

One way out of the bind is the common practice of vocabulary instruction; that is, we ought to, and in fact often do, preteach the difficult vocabulary in a passage, emphasizing both word meaning and word identification. While vocabulary instruction is frequently recommended (e.g., Karlin, 1967; Durkin, 1974), the method is often vague or unspecified. Sometimes teacher instruction is urged (McCullough, 1957; Gray and Holmes, 1938), and sometimes self-instruction through the use of a glossary (Haus, 1963). Research is unclear regarding the direct effect of specific vocabulary instruction, but a recent study (Tuinman and Brady, 1973) suggests that it may have a generalized effect on vocabulary test scores but not on comprehension test scores. What we need to know is whether or not students learn the vocabulary taught for a specific selection and whether such learning has a positive effect on comprehension of selections in which this vocabulary is embedded.

The present study was designed to investigate the relationship between vocabulary instruction and the comprehension of passages differing in the average word frequency level. We reasoned that vocabulary instruction might have an effect in versions employing unfamiliar (infrequent) words but not in versions employing familiar (frequent) words.

In order to more precisely specify the conditions under which these effects might operate, we included three other variables in the experiment. First, we manipulated content type, employing one scientific exposition and one historical narrative written in the biographical genre. Second, we varied question type over three levels varying from almost total textual reliance to minimal textual reliance, reasoning that vocabulary instruction might be more effective for those questions which were most textually reliant. Finally we blocked subjects on sex, male and female, reasoning that sex might interact with content type and responsiveness to instruction.

For this last consideration, there is some evidence to suggest an interaction between content type and sex. Girls appear to have somewhat different reading interests than boys (Chui, 1973), and boys' comprehension scores appear to be more affected by the level of personal interest in the story (Asher and Markell, 1974). American males generally achieve higher scores than females when the item content is scientific, mechanical, business, practical affairs, or mathematical, while females achieve higher scores than males on verbal tests when the content is human relations or the arts and humanities (Coffman, 1961; Donlon and Angoff, 1973).

In terms of possible experimental effects, we hypothesized that irrespective of the main effects for word frequency or vocabulary instruction, there would be a vocabulary instruction X frequency interaction such that vocabulary instruction would increase comprehension in passages written with low but not high frequency words. Second, we expected to find a vocabulary instruction X question type interaction, indicating that vocabulary instruction would be especially helpful for questions which were most textually reliant. Finally we expected that sex would interact with content type in accordance with previous findings regarding male and female interests for different types of reading material.

Method

Design

There were three between-subject variables: sex (male or female) word frequency (high or low) and vocabulary instruction (none, teacher directed or glossary). Content type (technical exposition or historical narrative) and question type (literal, inferential, or scriptal) were within subject variables. Equal numbers of male and female students were randomly assigned to each of the six word frequency X vocabulary instruction treatment conditions. All students answered five of each of the three types of questions following each of the two types of content. The data were analyzed using a three-between, two-within analysis of variance (Winer, 1971). Effects significant beyond $\alpha = .01$ were regarded as significant. This level was chosen because with such a large number of F tests, we were more concerned with Type I than Type II error.

Subjects

Subjects were 120 sixth grade students attending two elementary schools in a middle class suburb of Minneapolis. The subjects were selected by

participating teachers after being given instructions to choose sixth graders who read at or above grade level (on the basis of standardized test scores). Initially, 145 students were selected. From that group, six treatment groups of 20 students each were randomly selected, with the provision that each group contained 10 males and 10 females.

Materials

Content Type. Two modified stories (also used in the Marks et al. study) selected from the SRA Reading Laboratories, served as the experimental reading materials. "Conductor Moses" is a narrative, biographical account of Harriet Tubman's attempts to rescue slaves. "Bail Out" is a technical account of a pilot forced to bail out of his jet after it caught fire.

An SFI index of 50 (Carroll, Davies, and Richman, 1971) was used as the cut off point between high and low frequency words.[1] The stories were divided into blocks of six words. In approximately one-half the blocks, a high frequency synonym was substituted for a low frequency word of equal length; in the other half, a low for a high frequency word.[2] By grouping together the resulting high and low frequency blocks, two versions of each story were created differing only in the frequency of one-sixth of the words.

Questions. For each story, 15 multiple choice questions, each with four choices, were generated, five for each question type. Literal questions were generated by performing a wh-transformation on a specific text segment so as to replace an immediate constituent of the segment with the wh-word (Bormuth, Manning, Carr and Pearson, 1970). For example, "Harriet Tubman helped the slaves escape to Canada" $\Rightarrow$ "Where did Harriet Tubman help the slaves escape to?"

Text inference (inference) questions were generated by identifying a relationship between two grammatically unconnected text segments and asking a question which required their logical or pragmatic integration. For example, the text segment, "They walked close enough together so that they could touch each other in the dark. It wouldn't do for anyone to call out if he got lost from the party," might generate the text inference question, "Why did they walk close enough together so that they could touch each other in the dark." In general for text inference questions, both the question and the answer are in the text but the question-answer relationship is not obvious or grammatically cued.

Scriptal questions were even less textually reliant. Intuitively, they require a reading of the text to get into the right mental set (script (Schank, 1973) or schema (Anderson, 1977) if you prefer) to answer the question, but the answer must come from prior knowledge. It is just not available in the text. For example, the question, "When did Harriet Tubman live?" could be answered from a knowledge of American history and the Civil War. Nowhere in the text is even an approximate date given. In short, the question "came from" the text, but the answer had to come from prior knowledge.

One other important feature of questions. Each was written in such a way that at least one word in the item was a third synonym substitute for a word that been varied between the two versions (high and low frequency) of each story. For example if "afraid" appeared in the high frequency version and "apprehensive" in the low frequency version, then "anxious" appeared in the question. These so-called neutral synonyms were selected such that they possessed SFI values midway between the SFI values of the high and low frequency synonyms. This was done in order to

make certain that the question probes tapped text segments which differed between versions and to eliminate the likelihood that the questions were in any way biased toward one version or the other.

Once all the questions had been generated, a single random order for each set of 15 questions was determined and used for all subjects in all treatment conditions. The reliability of the total test (KR20) was .75, thus falling in between Downie's (1967) recommendations for standardized tests (.90) and classroom tests (.50).

Glossaries. Since one of the vocabulary pre-instruction treatments involved the use of a glossary, we created two glossaries (one high frequency and one low frequency) for each story. Each glossary contained 15 vocabulary items that seemed essential to an understanding of the story and were directly related to the questions.

In the glossary, each word was presented, defined and used in a sample sentence. For example: "Afraid. Afraid means anxious or apprehensive. For example, the girl was afraid of flunking the test." This is an example from the high frequency glossary. For the student reading the low frequency glossary, the same example read as follows: "Apprehensive. Apprehensive means anxious or afraid. For example, the girl was apprehensive about flunking the test." Each version of the glossary taught from the high frequency or low frequency version to the other two synonyms. In other words, the high frequency version presented the neutral (in the question) and low frequency synonyms, while the low frequency version presented the neutral and high frequency synonyms.

Procedure

Students were tested in groups of twenty (ten males and ten females), for a thirty minute period. The first two groups (one high and one low

frequency) received no vacabulary instruction. They were given a simple explanation of the purpose of the study and handed the first story to read. Each child, when finished with the story, raised his or her hand, returned the story to the experimenter, and received the fifteen multiple choice questions for that story. The same procedure was repeated with the second story. Story order was randomized within each treatment group.

The second two groups received vocabulary instruction in the form of the hand-out glossary. The experimenter handed each student a one page glossary along with each story, with the following instructions: "You will probably want to read the definitions before you read the story." The third two groups received direct vocabulary instruction by the experimenter before they read each story. They were told by the experimenter: "I'm going to go over some words in the story which might give you trouble. This will probably help you answer the questions at the end of the story." The experimenter then orally reviewed the same words on the previous glossary list, presenting exactly the same definition for each word. The difference was that this time the children heard each definition, but did not have the opportunity to read it. Under no conditions did the students have the opportunity to refer to the text while answering questions.

Results

Of the five main effects, only those associated with the within-subjects variables reached significance. The content type effect, $F(1,108) = 46.513$ $p < .01$, indicated that the scientific exposition ($\bar{x} = 10.45$) was more difficult than the historical narrative ($\bar{x} = 11.98$). There was considerably variation among question types, with scriptal questions proving

the easiest ($\bar{x}$ = 8.28), followed by literal ($\bar{x}$ = 7.75) and text inference questions ($\bar{x}$ = 6.40), F (2,216) = 70.34, $p < .01$. However, both of these variables were involved in higher order interactions.

The lack of a student sex effect was masked by a cross over interaction between sex and content type, F (1,108) = 9.73, $p < .01$. Girls ($\bar{x}$ = 12.28) surpassed boys ($\bar{x}$ = 11.67) on the historical narrative, while boys ($\bar{x}$ = 10.82) surpassed girls ($\bar{x}$ = 10.08) on the technical exposition.

Furthermore, the two significant main effects, content type and question type, also interacted, F (2,216) = 23.35, $p < .01$. The interaction appears, in large measure, to be due to the differential effect of literal questions between content types. Whereas the difference between inference questions from one version to another ($\bar{x}$ diff = .28) is of the same magnitude as the difference for scriptal questions ($\bar{x}$ diff = .14), the same difference for literal questions was much larger ($\bar{x}$ diff = 1.13). In short, the technical exposition was a bit harder than the historical narrative for scriptal and text-inference questions, but it proved much harder for the literal questions. In fact most of the main effect of content type is attributable to the particularly depressing effect of literal questions for the technical exposition.

Of the sixteen three and four way interaction tests, one proved to be significant: vocabulary x content x question type, F (4,216) = 3.11, $p < .01$. The sheer power of a test with that many degrees of freedom should elicit caution in interpreting the interaction. Interestingly it is the only point in the statistical analysis where vocabulary instruction emerged as a significant factor.

About all the interaction analysis merits is to say that relative to no instruction or glossary instruction, teacher instruction was particularly facilitating for inference questions in the technical exposition. In fact, if one compares the nine comparable question type X vocabulary instruction means across the two content types, there is only one case in which a technical exposition mean was greater than an historical narrative mean, i.e. for teacher instruction ($\bar{x}_{HN} = 2.95$, $\bar{x}_{TE} = 3.43$).

Discussion

Few of the experimental hypotheses were verified. First, we were not able to replicate the Marks, Doctorow and Wittrock findings. High frequency versions ($\bar{x} = 11.34$) were no more difficult than low frequency versions ($\bar{x} = 11.07$), $F(1,108) = .510$, $p. \geq .5$. Granted we crossed frequency with sex and vocabulary instruction between subjects. Yet none of the two or three way interactions involving those three variables approached significance (the largest F ratio was 1.74 for the three way interaction). Furthermore, looking at the simple effects within the no instruction condition summed across sex, there is no hint of an effect for frequency. The only systematic difference between the Marks et al. study and the present study was question type. Questions in the present study were written in a neutral language based upon text segments containing words whose lexical form varied between high and low frequency versions. Apparently this neutrality was sufficient to destroy the frequency effect. However, we have perused the questions used by Marks et al. and we find no systematic bias that could account for such variant findings. Even so, the results of the present study are congruent with a long line of research suggesting that frequency manipulations by themselves are not sufficiently powerful to affect comprehension (Klare, 1974-75).

Second, we were not able to demonstrate any effect for vocabulary instruction at all, let alone one exclusively for low frequency stories or particular question types. Here, however, we are forced to admit to a methodological weakness: the vocabulary instruction in this study was neither intensive, systematic nor complete. Nor did we assess prior knowledge. It is possible that these students (sixth grade students reading at grade level) knew all the critical low frequency words, although this is unlikely in view of the 10-11 grade level readability values. The only shred of support for vocabulary intervention comes from the content type X question type X vocabulary instruction interaction. It seems to say that if you have technical material and difficult text integration questions, it helps to have a teacher present the words critical for understanding the questions. What is needed is a stronger test of vocabulary instruction--some intensive concept building--in situations where one is sure the concepts are unfamiliar. On the other hand, the weak tests of vocabulary intervention in the present study may be more typical of what actual happens in classrooms.

The sex X content interaction is consistent with the research on sex differences and content (Coffman, 1961; Donlon and Anghoff, 1973), suggesting that boys are more interested in technical matters while girls are more interested in family relationships. However, the present data are not congruent with Asher and Markell's (1974) conclusion that girls are more tolerant of shifts away from personally interesting material.

The question type variable was originally included in order to determine whether or not some of the more central variables produced effects generalizable across different kinds of comprehension probes. It turned out to be the most potent variable in the study, and the results are somewhat counter-intuitive.

As educators we tend to regard factual recognition items as representing the lowest (and hence the easiest) level of comprehension (cf. Bloom, 1956; Barrett, 1968). Yet scriptal questions, which require logical connections between textual information and information stored in memory, were at least as easy as literal questions, and they were definitely easier for the technical exposition (i.e., the content type X question type interaction).

Perhaps more surprising is the low level of performance for text inference questions. Remember these questions required the reader to make a logical or pragmatic (statistical or probabilistic) connection between two textually presented segments of text (propositions).

In relationship to what gets into long term memory storage the question is, Why should it be more difficult to process two textual propositions and draw some connection between them than it is to process and store a single textual proposition (literal) or to relate a single text proposition to a proposition already in long term storage (scriptal)?

If one hypothesizes that memory for propositions is a function of how well they fit with pre-existing schema (e.g. Anderson, 1977, or Bower, 1976) the present data are more reasonable. If schematic knowledge is strong, as apparently it was for things like symptoms of periods of history or most likely motivations for human actions (the kinds of content tapped by scriptal questions), propositions are easily assimilated into memory and readily available for retrieval. This is reflected in the higher scores for scriptal questions. When schematic knowledge is weak (as it probably was for the very specific factual propositions in both the literal and text inference items), propositions are less likely to remain in storage. Furthermore, when the relationship between two propositions is not explicit (e.g. in the text inference questions) it is possible that the two proposi-

tions get into memory storage but they get into different schemata. Hence, retrieval of a relationship between them is more difficult.

If this explanation is accurate, then comprehension for text inference questions should be improved relative to comprehension for literal and scriptal questions if subjects are allowed free access to the text while answering questions, i.e. when subjects are allowed to actively search the text for such relationships. There should be almost no improvement for scriptal questions since they are as reliant on schemata as they are on text. Comprehension for literal questions should be improved by an intermediate amount.

As a follow-up, we gave the low frequency--no vocabulary instruction--versions of the two stories to a group of fourteen subjects from the same population, allowing them free access to the text while answering the questions. We contrasted this new group (12 of 14 were girls) with the comparable no access--low frequency--no vocabulary instruction female group (N=10). We found partial support for this explanation. The difference between access and no access groups was significant for inference questions ($\bar{x}_{diff} = 2.37$), $t(22) = 4.23$, $p. < .01$, but not significant for literal ($\bar{x}_{diff} = .51$) or scriptal ($\bar{x}_{diff} = .5$) questions.

While such an explanation is largely conjecture, it is loosely consistent with the relative difficulty of question types in the main study and quite consistent with the findings in the follow-up study. And the notion of schematic assimilation continues to gather empirical support (Anderson, 1976; Bower, 1976; Thorndyke, 1977). Unfortunately, schematic knowledge was not assessed or manipulated independently in the present investigation. Hence the conjecture label must stand until more appropriate research is conducted.

The limitations of this study deserve note. First, a stronger test of the vocabulary instruction variable could have been provided. Second, even though the passages were lengthy, one passage each of the two content types hardly constitutes a sample from some population of discourse one might like to generalize to. Third, inferences about content type are limited by the fact that protagonist sex, topic, and style were completely confounded. One was a historical narrative with a female protagonist; the other, a scientific exposition with a male protagonist. The confounding was not accidental. We chose to maximize the likelihood that content type would operate as an interactive factor. It remains for future research to sort out the relative impact of the three variables. Fourth, we chose a rather restricted population of subjects. With less able readers, it is not unreasonable to expect shifts in performance across question types or frequency levels.

Perhaps the most important contribution of the present study is that it establishes the necessity of examining comprehension as a multi-level factor incorporating both mind and (reading) matter.

Footnotes

1. The SFI scale is logarithmic: SFI = 40 corresponds to 100 occurrences per million, SFI = 50, 10; SFI = 60, 1. It was not altogether clear what Marks, Doctorow and Wittrock had done; however by examining a sample of 70 substitutions, we inferred that they must have used SFI = 50 as a cut-off.

2. We tried to parallel Marks et al. as closely as possible. They indicated that they tried to substitute "the word most amenable to substitution." We adopted a rule that we would try to substitute "the first lexical token for which we could find a synonym that met the high/low frequency criterion." In practice, our rule did not differ significantly from that used by Marks et al.

References

Anderson, Richard. The notion of schemata and the educational enterprise. In R. Anderson, J. Spiro, and W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Lawrence Erlbaum, Associates, 1977.

Asher, S. R. & Markell, R. A. Sex differences in comprehension of high and low interest reading material. Journal of Educational Psychology, 1974, 66, 680-687.

Barrett, Thomas. Taxonomy of cognitive and affective dimensions of reading comprehension. Unpublished paper. In Helen Robinson (Ed.), Innovation and Change in Reading Instruction, NSSE Yearbook, v. 27. Chicago: University of Chicago Press, 1968.

Bloom, Benjamin et al. Taxonomy of educational objectives, handbook I, cognitive domain. New York: Longmans, Green and Co., 1956.

Barmouth, John R., Julian Carr, John Manning and David Pearson. Children comprehension of between-and within-sentence syntactic structures. Journal of Educational Psychology, 1970, v. 61, 5, 349-357.

Carroll, J. B., Davies, P., and Richman, B. Word frequency book. New York: American Heritage Publishing Co., Inc., 1971.

Chiu, L. Reading preferences of fourth grade children related to sex and reading ability. The Journal of Educational Research, 1976, 66, 369-373.

Coffman, W. E. Sex differences in responses to items in an aptitude test. Eighteenth Yearbook, National Council on Measurement in Education, Edith M. Huddleston, (Ed.), 117-24. Ames, Iowa: National Council on Measurement in Education, 1961.

Dale, E. and Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20, 37-54.

Donlon, T. F. and Angoff, W. H. The scholastic aptitude test. The College Board Admissions Testing Program. W. H. Angoff, (Ed.), 15-47. New York: College Entrance Examination Board, 1971.

Downie, N. M. Fundamentals of measurement: techniques and practices (2nd ed.). New York: Oxford University Press, 1967.

Durkin, D. Teaching them to read (3rd ed.). Boston: Allyn and Bacon, Inc., 1974.

Fry, E. G. A readability formula that saves time. Journal of Reading, 1968, 11, 513-516, 575-578.

Gray, W. S. and Holmes, E. The development of meaning vocabularies in reading. Chicago: The University of Chicago Press, 1938.

Howes, D. H. and Solomon, R. L. Visual duration threshold as a function of word probability. Journal of Experimental Psychology, 1951, 41, 401-410.

Huus, H. Reading. Perspectives on elementary reading: principles and strategies of teaching, R. Karlin (ed.). New York: Harcourt Brace Jovanovich, Inc., 1973.

Karlin, R. Research results and classroom practices. The Reading Teacher, 1967, 21, 211-216.

Klare, George R. Assessing Readability. Reading Research Quarterly, 1974-75, X, 62-102.

Marks, C. B., Doctorow, M. J., and Wittrock, M. C. Word frequency and reading comprehension. Journal of Educational Research, 1974, 67, 259-262.

McCullough, C. What does research reveal about practices in teaching reading? English Journal, 1957, 66, 475-490.

Pearson, P. D. and Studt, A. Effects of word frequency and contextual richness on children's word identification abilities. Journal of Educational Psychology, 1975, 67, 89-95.

Schank, Roger, C. Identification of conceptualizations underlying natural language. In. R. C. Schank and K. M. Colby (Eds.), Computer Models of Thought and Language. San Francisco: Freeman, 1973.

Solomon, R. L. and Postman, L. Frequency of usage as a determinant of recognition thresholds for words. Journal of Experimental Psychology, 1952, 43, 195-201.

Thorndyke, Perry. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Tuinman, J. J. and Brady, M. E. How does vocabulary account for variance on reading comprehension tests? A preliminary to an instructional analysis. Interaction: research and practice for college-adult reading, P. L. Nacke (ed.), twenty-third year book of the National Reading Conference. Clemson, S. C.: National Reading Conference, 1974, 176-184.

Winer, B. J. Statistical principles in experimental design, 2nd ed. New York: McGraw Hill, 1971.

Zipf, George. The meaning-frequency relationship of words. Journal of General Psychology, 1945, 33, 251-256.